REMARKS:

This paper is herewith filed in response to the Examiner's Office Action mailed on January 25, 2008 for the above-captioned U.S. Patent Application. This office action is a rejection of claims 1-39 of the application.

More specifically, the Examiner has rejected claims 1-39 under 35 USC 103(a) as being unpatentable over Brecher (US7,054,754) in view of Shanahan. The Applicants respectfully traverse the rejections.

Regarding the rejection of claim 1 under 35 USC 103(a) the Applicants respectfully disagree with the rejection.

In the Office Action the Examiner states:

> "[Brecher discloses] determining that a first token considered of the plurality of tokens comprises a chemical name fragment (naphthoxy and phenacyl; column 12, lines 10-33), wherein determining comprises: <u>examining syntax of the first token (scanning for syntactic significance; column 3, lines 40-60 and column 8, lines 4-48)</u>, and taking into account the syntax (scanning fnr syntactic significance; column 3, lines 40-60 and column 8, lines 4-48) and the context (context; column 3, lines 14-60 and column 11, lines 22-42)," (emphasis added).

Claim 1 recites in relevant part:

> "determining that a first token considered of the plurality of tokens comprises a chemical name fragment, wherein determining comprises: examining syntax of the first token, examining context of the first token with respect to at least one adjacent token of the plurality of tokens."

The Applicants submit that Brecher does not relate to determining that a token comprises a chemical name fragment where determining comprises examining syntax of the token.

As cited Brecher discloses:

"The name is scanned from left to right and is copied, possibly with changes as now described, into a new temporary buffer (step 2020). During scanning, open- and close-parentheses and other enclosing marks are counted, and depths of enclosing marks are monitored. With some exceptions, characters are copied to the new buffer unmodified. Commas that are not enclosed within any level of enclosing marks are not copied, but are instead converted to @ signs. For simplicity, any space characters or additional commas immediately following such a comma are treated as having no syntactic significance, and are not copied.," (emphasis added), (col. 3, lines 40-50).

The Applicant submits that here the Examiner has not shown where Brecher is determining a token comprises a **chemical** name fragment by examining the syntax of the token. Brecher as cited by the Examiner appears to relate to syntactic significance of spaces and additional commas immediately following a comma. Further, it is noted that in each case Brecher indicates that the spaces and additional commas "are treated as having **no syntactic significance**."

Further, as cited Brecher discloses:

"The recognition of parentheses and other enclosing marks, if any, is integral to the name fragmentation process. During the fragmentation, the phrase surrounded by the innermost pair of enclosing marks is parsed as a unit, and is then consolidated as a unit according to a consolidation process described below with respect to the full name. Accordingly, each group within a set of enclosing marks is treated as a single unit, which is consistent with the syntactic meaning of enclosing marks," (emphasis added), (col. 8, lines 19-27).

At Col. 8 lines 25-27, Brecher describes that treating a phrase with enclosing marks such as a parenthesis is consistent with the syntactic meaning of the marks. However, this is in Brechers name fragmentation process (col. 8, lines 19-20). **Brecher does not use those marks or any other syntax to determine if what is within those marks is a chemical name fragment as opposed to some other fragment type**. The Applicants submit that here Brecher appears to disclose that a phrase surrounded by a pair of enclosing marks is **parsed as a unit** based on the enclosing marks. The Applicants submit that this method in Brecher is clearly distinguishable from claim 1 at least for the reason that syntax of a first token is examined after a text document has already been partitioned into a plurality of tokens and for each considered token there is a

step of examining the syntax of that token to determine that it comprises a chemical name fragment. As cited by the Examiner Brecher relates to grouping or consolidating a phrase into a single unit based on a perceived syntactic meaning of marks that surround the phrase. It appears that in Brecher the only syntactic meaning to be considered is in regards to enclosing marks for the purposes of name fragmentation. In claim 1, syntax of the first token is examined to determine that it comprises a **chemical** name fragment (as opposed to some other fragment), and that syntax is examined necessarily AFTER the text is partitioned into tokens.

The Applicants can not find in all of Brecher where there can be seen to be disclosed or suggested determining that a first token considered of a plurality of tokens already partitioned from a text document comprises a chemical name fragment where determining comprises examining the syntax of **the first token** as in claim 1.

Further, in the rejection of claim 1 the Examiner states:

> "applying a plurality of regular expressions (regular expression; column 5, lines 41-45), rules (rules; column 2, lines 59-65) and a plurality of dictionaries to recognize chemical name fragments (dictionary; column 6, lines 60-67), comprised of a prefix dictionary (prefix; column 9, line 52 — column 10, line 27) and a suffix dictionary (suffix; column 11, lines 43-59)," emphasis added).

As cited Brecher discloses:

> "In a specific embodiment, a fragment is determined to be meaningful ("recognized") if an exact match for the fragment is found in a dictionary of known text strings ("lexicon") that is maintained by the system," (emphasis added), (col. 6, lines 36-39).

> Each known text string is associated in the lexicon with at least one data object known as a nomToken (FIG. 6). A nomToken includes the text of the known text string as its name and is described by Type and Subtype data members, which allow similar fragments to be grouped in accordance with two levels of similarity," (emphasis added), (col. 6, lines 40-45).

The Applicants submit that Brecher as cited appears to merely disclose **a dictionary of known text strings (lexicon)**. Moreover, as stated above Brecher indicates that each known text string is

13

associated in the lexicon with at least one data object known as a **nomToken**.

Further, as cited Brecher discloses:

"<u>A nomToken of type kTypePrefix</u>, such as "pent" or "penta", may refer implicitly to an alkyl or heteroatomic chain. [and] In a different environment, when followed by a nomToken of kTypeRoot, "penta" indicates that the root structure should be repeated, and its original designation as kTypePrefix is retained for later handling," (emphasis added), (col. 9, lines 55-67); and

"<u>The list is examined for nomTokens of type kTypeSuffix</u>. Such a nomToken ("yl") is found, and is found to be preceded by a nomToken of type kTypeRoot, which results in a recognized environment," (emphasis added), (col. 11, lines 43-47).

The Applicants note that the Examiner appears to equate examining for different types of nomTokens in Brecher with a plurality of dictionaries comprised of a prefix dictionary, and a suffix dictionary to recognize chemical name fragments as in claim 1. As stated above, Brecher clearly discloses that "Each known text string is associated **in the lexicon** with at least one **data object known as a nomToken**." The Applicants submit that the lexicon of Brecher which apparently contains the at least one data object known as a nomToken can not be seen to relate to **a plurality of dictionaries** used to recognize chemical name fragments.

The Applicants contend that Brecher fails to disclose or suggest at least where claim 1 recites applying to the first token a plurality of regular expressions, rules, and **a plurality of dictionaries** comprised of a prefix dictionary, and a suffix dictionary to recognize chemical name fragments.

In addition, in the rejection of claim 1 the Examiner states:

[Brecher discloses] combining (concatenate) the first token with at least on the adjacent tokens (adjacent token) that are determined to be a chemical name fragment into a complete chemical name (column 8, lines 29-48), <u>but does not specifically teach assigning parts of speech. Shanahan discloses a method</u>

assigning the complete name with one part of speech and storing in a memory the complete chemical name assigned with the one part of speech (part-of-speech; column 10, lines 42-65), to denote the grammatical usage.

As cited Shanahan discloses:

Entities include proper names (e.g., people, places, organizations, etc.), times, locations, amounts, citations (e.g., book titles), addresses, etc. Entities can be recognized using a variety of known techniques that may include any one or a combination of regular expressions, lexicons, keywords, and rules. A lexicon is typically a database of tuples of the form <entity-string, part-of-speech-tag, entity-type> where: an entity-string is the string characters that make up the entity (e.g., a person's name "John Smith"); a part-of-speech-tag, which is optional, denotes the grammatical usage of the entity (e.g., as a noun, noun phrase, verb, etc.) [...]," (emphasis added); and

"Entities can be recognized by string matching or by using regular expressions. For example, a person's name could be recognized as two capitalized words. Regular expressions can be expressed in terms of the actual textual document content (i.e., words) or in terms of the linguistic markup associated with the textual content. This linguistic markup could include part of speech tags (such as noun phrases, nouns, etc.) or shallow parsing tags," (emphasis added), (col. 10, lines 42-65).

The Applicants submit that Shanahan as cited does not disclose or suggest at least where claim 1 recites assigning the **complete chemical name** with one part of speech and storing in a memory the complete chemical name assigned with the one part of speech.

As cited Shanahan appear to merely disclose that markup language which includes part of speech tags is associated with "proper names (e.g., people, places, organizations, etc)." The Applicants can not find in all of Shanahan where it is disclosed or suggested assigning a **complete chemical name** with one part of speech.

Furthermore, the Applicants note that Shanahan discloses:

"Initially, document service requests analyze a document by linguistically

processing the document to recognize entities within the document. These entities can be strings from a list (e.g., list of medicine names), or <u>regular expressions describing a multiplicity of entities</u> (e.g., a proper name recognizer, <u>a chemical formula recognizer</u>, etc.), or elements recognized by linguistic processing (e.g., noun phrases, words in a subject-verb relations, etc.)," (emphasis added), (col. 53, lines 10-16)

The Applicant submits that in a single instance, as stated above, Shanahan discloses a "chemical formula recognizer." However, the Applicants can not find in all of Shanahan where it is disclosed or suggested that a part of speech tag is associated with the chemical formula in Shanahan. Moreover, the Applicant contends that the mention of a **chemical formula** recognizer in Shanahan clearly can not be seen to relate to a **complete chemical name** as in claim 1. The Applicants submit that neither Brecher nor Shanahan can be seen to disclose or suggest **assigning a complete chemical name with one part of speech** as in claim 1.

The Applicants contend that for at least the reasons stated the references cited can not be seen to disclose or suggest claim 1 and the rejection of claim 1 should be removed.

In addition the Applicants note that independent claims 13, 25, and 35 recite features similar to claim 1 as stated above. Thus, for at least the reasons already stated the references cited can not be seen to disclose or suggest these claims.

Regarding the rejection of claim 2 the Applicants note that for at least the reasons already stated the references cited are not seen to disclose or suggest at least where claim 2 recites "where the **complete chemical name is assigned** a noun phrase **part of speech**."

Further, for at least the reason that claims 14, 26, and 36 recite features similar to claim 2 the references cited are not seen to disclose or suggest all of claims 2, 14, 26, and 36 and the rejections of these claims should be removed.

Regarding the rejection of claims 4, 16, and 28 the Examiner states:

"Regarding claims 4,16 and 28, Brecher discloses a method to process a document, but does not specifically teach where said plurality of dictionaries comprises a dictionary of stop words to eliminate erroneous chemical name fragments," and;

"Shanahan discloses a method where said plurality of dictionaries comprises a dictionary of stop words to eliminate erroneous chemical name fragments (stop words eliminated; column 27, lines 28-36 with column 37, lines 28-45 and column 49, lines 58-65), to discard un-important words," (emphasis added).

The Applicants contend that the method in Shanahan clearly can not be seen to relate to **chemical name fragments**. The Applicants submit that although Shanahan mentions a chemical formula recognizer one time, Shanahan clearly fails to disclose or suggest any operation regarding a chemical name fragment. Moreover, the Applicants contend that no where in Shanahan is there found any disclosure or suggestion of an erroneous chemical name fragment or even a chemical name fragment. The Applicants submit that citing Shanahan in the rejections as overcoming this admitted shortfall of Brecher is clearly unsupported and improper.

In addition, regarding the rejection of claim 4 the Applicants contend that for at least the reasons already stated Shanahan can not be seen to disclose or suggest at least where claim 4 recites "where said plurality of dictionaries further comprise a dictionary of stop words to eliminate **erroneous chemical name fragments.**"

Furthermore, regarding the rejection of claim 16, for at least the reasons already stated, the Applicants contend that the references cited can not be seen to disclose or suggest at least where claim 16 recites "where said plurality of dictionaries further comprise a dictionary of stop words to eliminate **erroneous chemical name fragments.**"

Further, in regards to the rejection of claim 28 the Applicants contend that for at least the reasons already stated the references cited are not seen to disclose or suggest at least where claim 28 recites "where said plurality of dictionaries further comprise a dictionary of stop words to eliminate **erroneous chemical name fragments.**"

In addition, regarding the rejections of claims 5, 17, and 29 the Examiner states:

> "Regarding claims 5, 17 and 29, Brecher discloses a method to process a document, but does not specifically teach filtering recognized chemical name fragments using a list of stop words to eliminate erroneous chemical name fragments. Shanahan discloses a method comprising filtering recognized chemical name fragments using a list of stop words to eliminate erroneous chemical name fragments (stop words eliminated; column 27, lines 28-36 with column 37, lines 28-45 and column 49, lines 58-65), to discard un-important words," (emphasis added).

The Applicants note that again the Examiner indicates that Shanahan is related to recognizing chemical name fragments. For at least the reasons already stated the Applicants again assert that citing Shanahan in the rejections in order to overcoming this admitted shortfall of Brecher is clearly unsupported and improper. Thus, the rejections of all claims 5, 17, and 29 should be removed.

Furthermore, the Applicants request that the Examiner provide reference to support in Shanahan relating to recognizing chemical name **fragments** in a non-final Office Action or allow the claims 4-5, 16-17, and 28-29.

Further, in the rejection of claims 38 and 39 the Examiner states:

> "Regarding claims 38-39, Brecher discloses a method and computer program product where identifying tokens to be ignored comprises applying a negative dictionary (list of tokens "mg/ml") to the plurality of tokens (column 8, lines 4-61) and wherein the plurality of dictionaries **consists** of the prefix dictionary (prefix; column 9, line 52 - column 10, line 27), the suffix dictionary (suffix; column 11, lines 43-59), and the negative dictionary (list of tokens; column 8, lines 4-61)," (emphasis added).

The Applicants contend that Brecher can not be seen to disclose or suggest at least where claims 38-39 similarly recite where identifying tokens to be ignored comprises applying a negative dictionary to the plurality of tokens and **wherein the plurality of dictionaries <u>consists</u> of the prefix dictionary, the suffix dictionary, and the negative dictionary.**

Firstly, the Applicants respectfully submit that as stated above Brecher appears to disclose **only one dictionary**. The dictionary referred to is a "dictionary of known text strings ("lexi-con")," (col. 6, lines 36-40).

Moreover, the Applicants note that as stated above in regards to claim 1 the Examiner appears to equate nomTokens types in the lexicon of Brecher with each of a plurality of dictionaries. The Applicants contend that even if this were proper, which is not agreed with, **the multiple nomTokens identified in Brecher greatly exceeds the three dictionaries listed in claims 38 and 39**. The Applicants contend that Brecher clearly can not be seen to disclose or suggest at least where claims 38 and 39 recite "**wherein the plurality of dictionaries <u>consists</u> of the prefix dictionary, the suffix dictionary, and the negative dictionary**."

Further, the Applicants' representative notes that a perceived lack of support for the rejection of claims 38 and 39 was relayed to the Examiner, without resolve, in a telephone conversation on April 9, 2008. See MPEP 211.03 and 2173.05(h) for the proposition that the transition phrase "consisting of" excludes unrelated elements. If the various nomTokens of Brecher are separate dictionaries as the rejection of claim 1 implies, then Brecher is inoperative if it were restricted to only the three dictionaries recited in claims 38 and 39.

Respectfully, it is requested that the Examiner reconsider the rejections of claims 38 and 39, and remove the rejections.

The Applicants submit that for at least the reasons stated the combination of Brecher and Shanahan, though not agreed to as proper, would still not disclose or suggest the present invention.
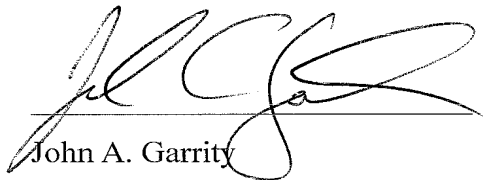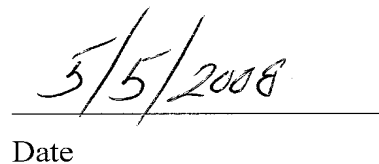
Further, for at least the reason that the claims 2-12; 14-24; 26-34; and 36-37 depend from claims 1, 13, 25, and 35 respectively, the references cited are not seen to disclose or suggest all claims 1-39.

Based on the above explanations and arguments, it is clear that the references cited cannot be seen to disclose or suggest claims 1-39. The Examiner is respectfully requested to reconsider and remove the rejections of claims 1-39 and to allow all of the pending claims 1-39 as now presented for examination.

For all of the foregoing reasons, it is respectfully submitted that all of the claims now present in the application are clearly novel and patentable over the prior art of record. Should any unresolved issue remain, the Examiner is invited to call Applicants' attorney at the telephone number indicated below.

Respectfully submitted:

John A. Garrity
Date 5/5/2008

Reg. No.: 60,470

Customer No.: 48237

HARRINGTON & SMITH, PC

4 Research Drive

Shelton, CT 06484-6212

Telephone: (203)925-9400

Facsimile: (203)944-0245

email: jgarrity@hspatent.com